

Understanding Society
Working Paper Series

No. 2013 – 06

November 2013

***Understanding Society* Innovation Panel Wave 5:
Results from Methodological Experiments**

Jonathan Burton (ed.)

Contributors: Katrin Auspurg¹, Jonathan Burton², Carl Cullinane³, Adeline Delavande², Laura Fumagelli², Maria Iacovou⁴, Annette Jäckle², Olena Kaminska², Peter Lynn², Paul Mathews², Gerry Nicolaas³, Cheti Nicoletti⁵, Cong Ye^{6,7}, Basit Zafar⁷

¹ University of Konstanz

² Institute for Social and Economic Research, University of Essex

³ NatCen Social Research, London

⁴ University of Cambridge

⁵ University of York

⁶ American Institutes for Research (AIR)

⁷ Research and Statistics Group, Federal Reserve Bank of New York

***Understanding Society Innovation Panel Wave 5:
Results from Methodological Experiments***

Jonathan Burton (ed.)

Contributors: Katrin Auspurg¹, Jonathan Burton², Carl Cullinane³, Adeline Delavande², Laura Fumagelli², Maria Iacovou⁴, Annette Jäckle², Olena Kaminska², Peter Lynn², Paul Mathews², Gerry Nicolaas³, Cheti Nicoletti⁵, Cong Ye⁶, Basit Zafar⁷

¹ University of Konstanz

² Institute for Social and Economic Research, University of Essex

³ NatCen Social Research, London

⁴ University of Cambridge

⁵ University of York

⁶ American Institutes for Research (AIR)

⁷ Research and Statistics Group, Federal Reserve Bank of New York

Non-technical summary

The *Understanding Society* survey includes what is known as an ‘Innovation Panel’ sample (IP). This sample of originally 1500 households is used to test different methods for conducting longitudinal surveys in order to produce the highest quality data. The results from the Innovation Panel provide evidence about the best way to conduct a longitudinal survey which is of relevance for all survey practitioners as well as influencing decisions made about how to conduct *Understanding Society*. This paper reports the experiments with the mixed-mode design and early results of the methodological tests carried out at wave 5 of the Innovation Panel in the spring of 2012.

IP5 was the second mixed-mode wave of the Innovation Panel. IP2 had experimented with telephone interviewing in addition to face-to-face personal interviewing. IP5 uses a design in which a random set of households are allocated to a sequential mixed-mode design. The adults in these households were first approached by letter and email where possible and asked to complete their interview on-line. Those who did not respond on-line were then followed up by face-to-face interviewers. The remaining third of households were issued directly to face-to-face interviewers.

The methodological tests included an experiment testing the effects of different incentives offered to respondents in advance of fieldwork on response rates, the use of ‘persuasive’ text in advance letters and the day of mailing of the advance materials. Further tests within the interview examined subjective expectations about the returns to university education, the implicit comparison group when asking questions about satisfaction, the effect of different wording in questions employing dependent interviewing, asking about mode preference, using vignettes to measure partnership satisfaction with the division of housework and the context effects of close social network reporting on childbearing intentions.

Understanding Society Innovation Panel Wave 4:
Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Katrin Auspurg¹, Jonathan Burton², Carl Cullinane³, Adeline Delavande², Laura Fumagelli², Maria Iacovou⁴, Annette Jäckle², Olena Kaminska², Peter Lynn², Paul Mathews², Gerry Nicolaas³, Cheti Nicoletti⁵, Cong Ye⁶, Basit Zafar⁷

¹ University of Konstanz

² Institute for Social and Economic Research, University of Essex

³ NatCen Social Research, London

⁴ University of Cambridge

⁵ University of York

⁶ American Institutes for Research (AIR)

⁷ Research and Statistics Group, Federal Reserve Bank of New York

Abstract

This paper presents some preliminary findings from Wave 5 of the Innovation Panel (IP5) of *Understanding Society: The UK Household Longitudinal Study*. *Understanding Society* is a major panel survey in the UK. In February 2012, the fifth wave of the Innovation Panel went into the field. IP5 used a mixed-mode design, using on-line interviews and face-to-face interviews. This paper describes the design of IP5, the experiments carried and the preliminary findings from early analysis of the data.

Key words: longitudinal, survey methodology, experimental design, respondent incentives, questionnaire design.

JEL classification: C80, C81, C83

Contact: Jonathan Burton (jburton@essex.ac.uk) Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.

Contents

1	Introduction	2
2	<i>Understanding Society</i> : the UKHLS	2
3	Innovation Panel Wave 5: Design	3
	a. Call for experiments	4
	b. Sample	5
	c. Questionnaire design	5
	d. Response rates	7
4	Experimentation in IP5	11
	a. Unconditional Respondent Incentives (Peter Lynn)	11
	b. Subjective Expectations about the Returns to Schooling and the Decision to go to University (Adeline Delavande and Basit Zafar)	14
	c. Persuasion Messages in Advance Letter (Cong Ye).....	13
	d. Satisfaction and the implicit comparison group (Laura Fumagalli)	14
	e. Wording effects of dependent interviewing questions on the amount of change observed in panel data (Annette Jäckle)	18
	f. Tailoring mode of data collection in longitudinal studies (Olena Kaminska and Peter Lynn).....	22
	g. Weekday of Advance Mailing and Conditional Incentives Experiment (Carl Cullinane and Gerry Nicolaas).....	25
	h. Measuring Partnership Satisfaction with the Division of Housework (Katrin Auspurg, Maria Iacovou, Cheti Nicoletti)	27
	i. Family, Friends and Fertility: Context effects on childbearing intentions and close social network reporting (Paul Mathews).	32
	REFERENCES	38

1 Introduction

This paper presents early findings from the fifth wave of the Innovation Panel (IP5) of *Understanding Society*: The UK Household Longitudinal Study (UKHLS). *Understanding Society* is a major panel survey for the UK. The first three waves of data collection on the main sample have been completed, the fourth and fifth waves are currently in the field. Full Wave 1 and the Wave 2 data from the main samples are available from the UK Data Archive, with full Wave 3 data later in 2013. Data from a nurse visit to collected bio-markers from the general population sample are also available, with data from a nurse visit to BHPS sample members available later in 2013. Data for the first five waves of the Innovation Panel are also available from the UK Data Service¹.

One of the features of *Understanding Society*, alongside the large sample size (40,000 households at Wave 1), the ethnic minority boost sample and the collection of bio-markers, is the desire to be innovative. This has been a key element of the design of *Understanding Society* since it was first proposed. Part of this drive for innovation is embodied within the Innovation Panel (IP). This panel of almost 1500 households were first interviewed in the early months of 2008. The design in terms of the questionnaire content and sample following rules are modelled on *Understanding Society*. The IP is used for methodological testing and experimentation that would not be feasible on the main sample. The IP is used to test different fieldwork designs, new questions and new ways of asking existing questions.

The second wave of the Innovation Panel (IP2) was carried out in April-June 2009, the third wave (IP3) in April-June 2010 and the fourth wave in March-July 2011. The fourth wave of the Innovation Panel (IP4) included a refreshment sample of 465 responding households. Working Papers which cover the experimentation carried out in all four innovation panels are available from the *Understanding Society* website.² The data from the first five waves of the innovation panel are held at the UK Data Service. In March 2012, IP5 was fielded. This paper describes the design of IP5, the experiments carried and some preliminary findings from early analysis of the data. Section 2 outlines the main design features of *Understanding Society*. Section 3 describes the design and conduct of IP5. Section 4 then reports on the experiments carried at IP5. Section 5 concludes the paper by outlining plans for IP6.

2 *Understanding Society*: the UKHLS

Understanding Society is an initiative of the Economic and Social Research Council (ESRC) and is one of the major investments in social science in the UK. The study is managed by the Scientific Leadership Team (SLT), based at ISER at the University of Essex and including members from the University of Warwick and the Institute of Education. The fieldwork and delivery of the survey data is undertaken by NatCen Social Research (NatCen). *Understanding Society* aims to be the largest

¹ <http://discover.ukdataservice.ac.uk/series/?sn=2000053>

² <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2008-03>
<https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2010-04>
<https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2011-05>
<https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2012-06>

survey of its kind in the world. The sample covers the whole of the UK, including Northern Ireland and the Highlands and Islands of Scotland. *Understanding Society* provides high quality, longitudinal survey data for academic and policy research across different disciplines. The use of geo-coded linked data enables greater research on neighbourhood and area effects, whilst the introduction of bio-markers and physical measurements (Waves 2 and 3) opens up the survey to health analysts.

The design of the main-stage of *Understanding Society* is similar to that of the British Household Panel Survey (BHPS) and other national panels around the world. In the first wave of data collection, a sample of addresses was issued. Up to three dwelling units at each address were randomly selected, and then up to three households within each dwelling unit were randomly selected. Sample households were then contacted by NatCen interviewers and the membership of the household enumerated. Those aged 16 or over were eligible for a full adult interview, whilst those aged 10-15 were eligible for a youth self-completion. The adult interviews were conducted using computer-assisted personal interviewing (CAPI) using lap-tops running the questionnaire in Blaise software. Adults who participated in *Understanding Society* were also asked to complete a self-completion questionnaire, in which questions thought to be more sensitive were placed. The adult self-completions at Waves 1 and 2, and the youth self-completions, were paper questionnaires. From Wave 3 onwards the adult self-completion instrument was integrated into the Blaise instrument and the respondent used the interviewer's lap-top to complete that portion of the questionnaire themselves (Computer-Assisted Self-Interviewing, CASI).

In between each wave of data collection, sample members are sent a short report of early findings from the survey, and a confirmation-of-address slip, to allow them to confirm their address and contact details. Before each sample month is issued to field for a new wave, each adult is sent a letter which informs them about the new wave of a survey, includes a token of appreciation in the form of a gift voucher and also includes a change-of-address card. Interviewers then attempt to contact households and enumerate them, getting information of any new entrants into the household and the location of anyone who has moved from the household. New entrants are eligible for inclusion in the household. Those who move, within the UK, are traced and interviewed at their new address. Those people living with the sample member are also temporarily eligible for interview. More information about the sampling design of *Understanding Society* are available in Lynn (2009).³ From Wave 2, the BHPS sample has been incorporated into the *Understanding Society* sample. The BHPS sample are interviewed in the first year of each wave.

3 Innovation Panel Wave 5: Design

IP5 employed a mixed-mode design. This is not the first time a mixed-mode design was used on the IP, at IP2 a design which combined telephone and face-to-face interviewing was used. At IP5 the modes which were mixed were on-line (CAWI) and face-to-face (CAPI) interviewing.

A random selection of two-thirds of households were allocated to the mixed-mode design (WEB) with the remaining third of households allocated directly to face-to-face interviewers (F2F). The

³ <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2009-01.pdf>

fieldwork for the WEB group started two weeks earlier than the F2F fieldwork. Initially, advance letters were sent to adults in the WEB group which included a URL and a unique log-in code. Adults in the WEB group for whom we had an email address were also sent an email which included a link which could be clicked through to the web-site. There were two email reminders for adults with an email address who had not yet completed their interview on-line, sent three days apart. A reminder letter was then sent to all adults in the WEB group who had not completed their interview. This letter was sent just under two weeks after the initial advance letter.

At the end of two weeks, all adults who had not completed their interview were allocated to face-to-face interviewers. Adults who had started their interview on-line, but not reached the 'partial interview' marker, were issued to face-to-face interviewers. The interviewers were able to re-start the interview at the place at which the respondent had stopped. Also at this point the remaining third of households, those in the F2F group, were issued to interviewers. The two-week WEB-only period before face-to-face fieldwork was implemented so that the face-to-face interviewers would have their full allocation at the start of their fieldwork, rather than having non-responding WEB individuals being passed to them during the fieldwork period. This was done to allow the face-to-face interviewers to work more efficiently.

The WEB-only period ran from 11th May to 24th May, there was some variation because of a day-of-week experiment that was being implemented (see 4g below). The face-to-face fieldwork started 24th May and ran until 15th July. During this period the CAWI survey remained 'open' so that WEB individuals could complete their interview on-line during this fieldwork period. A re-issue period ran from 26th July to 30th August, during this period households which had not been contacted or had given 'soft' refusals were re-issued to interviewers. The CAWI survey was closed for this period. Interviewers reported that having the on-line survey open made it harder for them to persuade people to participate when they were contacted in person. It was reported that interviewers would be told by sample members that they would complete the survey on-line, rather than give a CAPI interview. By closing the CAWI survey, it was hoped that interviewers would then be able to persuade non-responding adults to complete their interview in-person.

Prior to the survey going into the field there were eight one-day briefings for the interviewers. The briefings were conducted by NatCen researchers, with staff from ISER contributing to provide information about the study and to talk in more detail about the experiments. The briefings were held in Brentwood, Bristol, Derby, Glasgow, Leeds, London (2 briefings) and Manchester. In total, 116 interviewers were briefed to work on IP5. Six interviewers also attended a de-brief session in Brentwood on 29th August 2012. The questionnaires used at IP5 are available from the *Understanding Society* website.⁴

a. Call for experiments

IP5 was the third time the Innovation Panel was open for researchers outside the scientific team of *Understanding Society* to propose experiments. A public call for proposals was made on 30th March 2011 with a deadline of 2nd June. Twenty-one proposals were received and twelve were accepted. Submissions came from within ISER (five), ISER in collaboration with other researchers (five) and

⁴ <https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires>

from outside ISER completely (eleven). Of those that were external to ISER, five were from institutions within the UK and five were from outside the UK. The twenty-one proposals were reviewed by a panel which included two ISER-based members of the *Understanding Society* scientific leadership team, and two members of the Methodology Advisory Committee to *Understanding Society* who were external to ISER. In addition to those experiments which were accepted through the public call, there were a number of core experiments which the Understanding Society senior leadership team wanted to run. These core experiments included the mixed-mode design and the main incentives experiment.

b. Sample

The sample issued for IP5 included the original sample and the refreshment sample which had been interviewed at IP4. The original sample at IP5 comprised those household who had responded at IP4, plus some households which had not responded at IP4. Households which had adamantly refused or were deemed to be mentally or physically incapable of giving an interview were withdrawn from the sample. There were 1,126 original sample households were issued at IP5; 966 were previous-wave responding households, 26 were non-contacts at IP4, 110 were refusals at IP4 and 24 were non-responding for other reasons. In addition, 489 households from the refreshment sample were also issued. In the refreshment sample it was only households that had participated at IP4 that were issued at IP5.

As discussed above, around two-thirds of the sample were allocated to the mixed-mode design in which sample members would be approached by letter and email (where possible) to complete their interview on-line. This experimental allocation covered both the original and refreshment sample. The table below shows the allocation to mode design by sample type.

Table 1: Allocation to mode design by sample type

	Original sample	Refreshment sample	Total
CAPI only (F2F)	374 33.2%	169 34.6%	543 33.6%
Mixed-mode (WEB+F2F)	752 66.8%	320 65.4%	1,072 66.4%
Total	1,126	489	1,615

c. Questionnaire design

The questionnaire at IP5 followed the standard format used in the previous Innovation Panels as well as the main-stage of *Understanding Society*. The interview included:

- Household roster and household questionnaire: 15 minutes per household
- Individual questionnaire: on average 31 minutes for each person aged 16 or over
- Adult self-completion: around 9 minutes, paper questionnaire or computer self-administered interview (CASI)

- Youth self-completion: 10 minutes for each child aged 10-15 years
- Proxy questionnaire: 10 minutes for adults ages 16 or over who are not able to be interviewed.

Unlike some previous IPs, IP5 did not include audio recording of any portions of the interview.

There were some changes made to the questionnaire to enable participants to complete it on-line. The wording of the question-specific interviewer instructions was reviewed. Where it was felt that the information would be useful to clarify the definition of the question, the text was re-worded and made part of the question. Where the information was useful in directing the participant how to answer the question, for example the format in which the answer was required, this was placed between the question and the response categories. In addition, the text which appeared on 'help' screens was reviewed and re-written to be more participant-focused rather than interviewer-focused.

In a CAPI interview, the interviewer is present to manage the process of collecting information. On *Understanding Society* the first thing the interviewer completes is the household enumeration grid. This instrument allows interviewers to record new entrants to the household and those who have left the household, and collect – or check – basic information on household members. Generally, the interviewer then administers the household questionnaire, which collects household-level information (such as rent or mortgage payments, utility bills, household consumption etc). This is done with one person in the household, preferably someone who will know the information required. Once this is complete, the interviewer then interviews each adult who is available and willing to be interviewed.

Moving to CAWI, the interviewer is no longer available to manage this process. It was necessary, therefore, to make the questionnaire suitable for self-completion by the sample member, and to enable the appropriate instrument to be available to the appropriate person. In the advance mailing, each adult was given a unique code to log in to their questionnaire. The first person in the household to log in would be asked to complete the household enumeration. It was felt that this could be done by anyone in the household aged 16 or older. A new question at the end of the enumeration grid asked who was the person or people responsible for paying household bills. A list of all the household members was presented and the participant could select one or more people. If the person completing the grid was not one of those people responsible, or their spouse/partner, they were directly routed to their own individual questionnaire. If the person completing the grid was an appropriate person, they were routed into the household questionnaire, and then their individual interview. If anyone else in the household logged in after the grid was complete, they would be directly routed to the household questionnaire if they were an appropriate person and it had not yet been done, or directly to their individual interview. Once the grid or the household questionnaire was complete, it did not come back on-line for anyone else in the household.

If a participant had started to answer their questionnaire and left the computer for 10 minutes, they were automatically logged out. This was for privacy and security reasons. The participant was able to log back in using the same process as they had originally logged in, and they would be taken to the place that they had left the interview. This also applies to those who had closed down the browser mid-interview. A 'partial interview' marker was put into place about two-thirds of the way

through the interview, after the benefits section. If a participant reached this stage, the interview was considered to be a ‘partial interview’. They could log back in and complete if they wanted, but otherwise they were not contacted by an interviewer. If the participant had not reached this marker before closing down the browser, they were sent an email overnight which thanked them for their work so far and encouraged them to complete the survey, giving them the URL to click through to the survey. Again, they would start at the point where they had left off. In addition, those who had started but not reached the partial interview marker were, after the initial two weeks, issued to face-to-face interviewers who would be able to finish the survey with them, from where they had left off.

d. Response rates

This section sets out the response rates for IP5 as a whole. Section 4f describes the effect of incentives on response rates. Table 1, below, sets out the response rates for eligible households for the refreshment sample and the original sample. It also separates out the response rate for households that had responded at IP4 and those that had not. These latter columns are for the original sample only, since non-responding refreshment sample households from IP4 were not issued to field at IP5.

Table 1: Household response at IP5

	Refreshment sample	IP5 original sample	IP4 responding	IP4 non-responding	Total
Responding	83.0% 401	75.5% 823	82.1% 769	35.3% 54	77.8% 1,224
Non-contact	6.2% 30	6.2% 67	3.8% 36	20.3% 31	6.2% 97
Refusals	9.1% 44	16.1% 175	12.3% 115	39.2% 60	13.9% 219
Other non-responding	1.7% 8	2.3% 25	1.8% 17	5.2% 8	2.1% 33
n	483	1,090	937	153	1,573

Excludes households which were ineligible at IP5.

For the original sample, three-quarters of households responded (75.5%). Households who had responded at IP4 were, not surprisingly, more likely to respond at IP5; just over four-fifths responding compared to just over one-third of IP4 non-responding households. Refreshment sample households were as likely to respond at IP5 as the original sample who had responded at IP4. The refusal rate in the refreshment sample was a little lower, and the non-contact rate a little higher, than the original sample, but these differences were not significant. The refreshment sample were allocated to groups that received a higher incentive (£10, £20, £30) than the original sample (£5, £10) and this may explain the relatively high response at their second wave in the study.

Table 2, below, shows the individual response rate within those households that were at least partially responding at IP5. Within households, there was a higher level of full interviews, and a

lower level of proxy interviews, for the refreshment sample compared to the original sample. Again, this may be linked to the generally higher incentive levels in the refreshment sample.

Table 2: Individual response within responding households at IP5

	Original sample	Refreshment sample	Total
Full interview	80.8%	85.3%	82.3%
	1,315	680	1,995
Proxy interview	7.2%	3.8%	6.1%
	117	30	147
Non-contact	4.9%	5.4%	5.1%
	80	43	123
Refusal	5.8%	3.8%	5.1%
	94	30	124
Other non-response	1.3%	1.8%	1.4%
	21	14	35
n	1,627	797	2,424

The table below shows the outcome at IP5 of those adults who participated at IP4. Over three-quarters of those who gave a full interview at IP4 also gave a full interview at IP5. Just over one-fifth of those who had a proxy interview taken on their behalf at IP4 also had a proxy interview at IP5 as well, although a little over a quarter gave a full interview at IP5. Around one-fifth of adults who were eligible for interview at IP4 but did not respond gave a full interview at IP5.

Table 3: Individual re-interview rates (whole sample)

IP5	IP4		
	Full interview	Proxy interview	Adult not interviewed
Full interview	77.2	26.5	22.7
Proxy interview	1.8	22.4	7.7
Refusal	1.0	9.4	11.0
Other non-respondent	3.7	9.0	8.9
Household not interviewed	14.7	30.9	47.2
Ineligible	1.6	1.8	2.5
N	2,219	223	598

At IP5 there was a mixed-mode design experiment. As described in section c above, two-thirds of households were allocated to a sequential mixed-mode design (CAWI followed by CAPI for non-respondents), whilst the remaining third were allocated to face-to-face interviewers. More information on the design of IP5 is available from the user manual for the data.⁵

In a mixed-mode survey design cost savings are potentially possible if an interviewer does not have to be sent to the household. On a household survey like Understanding Society this means that the

⁵ http://www.esds.ac.uk/doc/6849/mrdoc/pdf/6849_ip_waves_1to5_user_manual.pdf

household enumeration, household questionnaire and adult questionnaire for all those aged 16 or over in the household needs to be completed on-line. In a multi-person household, if one adult doesn't complete their interview on-line, an interviewer needs to visit the household in person. Table 4, below, shows the proportion of households fully responding by web. Tables 4, 5 and 6, as well as some of the description of these tables, are taken from the Understanding Society working paper "Going Online with a Face-to-Face Household Panel: Initial Results from an Experiment on the *Understanding Society* Innovation Panel", 2013-03 (Jäckle, Lynn, Burton).⁶

At IP5 almost a fifth of households in the original sample allocated to the mixed mode design were completed on-line (18.5%). This proportion was higher for those households who had participated at IP4 (21.0%) than non-responding IP4 households (4.5%). Over a third of households in the refreshment sample allocated to mixed modes were completed on-line (34.3%). This higher completion rate is explained by the incentive groups on the refreshment sample. For all households where individuals received the £10 incentive, the household completion rates were very similar; 24.7% for the original sample who had participated at IP4 and 23.0% for the refreshment sample (all of whom had participated at IP4).

Table 4: Proportion of Households Fully Responding by Web, by Sample and Unconditional Incentive

Proportion of households (<i>n</i>)	Original sample			Refreshment sample
	Total sample	IP4 respondent households	IP4 non-respondent households	Total sample
Total sample	18.5 (728)	21.0 (618)	4.5 (110)	34.3 (315)
Unconditional incentive:				
£5	17.5 (389)	19.5 (330)	6.8 (59)	—
£10	21.2 (311)	24.7 (263)	2.1 (48)	23.0 (87)
£20	—	—	—	37.1 (105)
£30	—	—	—	42.9 (111)
P	0.21	0.11	0.22	0.00

Note: The analysis by value of the unconditional incentive excludes $n=40$ split-off households. *P*-values from Wald tests of the equality of mean completion rates between incentive groups, adjusted for sample design.

The table below gives the household outcome at IP5 for each of the three sample types – the IP4 original sample responding households, the IP4 non-responding original sample households and the refreshment sample. The household response rate is given for the F2F and WEB samples in each of these groups.

Across all sample types the household response rate for the F2F sample was 78%, slightly higher than the household response rate for the mixed mode sample at 74.3%. However this difference was not statistically significant. For the original sample who had participated at IP4 and the refreshment sample, the F2F response was higher than the mixed-mode response, although not statistically significant. For the original sample who had not responded at IP4 the position was reversed, with the mixed-mode response rate being around 4 percentage points higher than the F2F sample (again, though, not statistically significant).

⁶ <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2013-03.pdf>

It does appear, however, that the mode design did affect the proportion of complete households; where all eligible adults participated. For the original sample who had participated at IP4, the proportion of complete household response was 63.9% in the F2F sample and 55.7% in the mixed-mode sample, this difference was statistically significant. For this sample type, the single F2F design elicited a higher non-contact rate but a lower refusal rate than the mixed-mode design.

Table 5: Household Response Rates

	Original sample									Refreshment sample		
	Total			IP4 responding			IP4 non-responding					
	F2F	MM	<i>P</i>	F2F	MM	<i>P</i>	F2F	MM	<i>P</i>	F2F	MM	<i>P</i>
HH response rate	78.0	74.3	0.22	84.1	81.1	0.29	32.6	36.4	0.66	85.1	81.9	0.45
Complete HHs	58.0	50.3	0.02	63.9	55.7	0.02	14.0	20.0	0.43	60.1	66.0	0.26
Partial HHs	20.1	24.0	0.13	20.2	25.4	0.07	18.6	16.4	0.72	25.0	15.9	0.01
Non-contact	6.9	5.8	0.49	5.6	2.9	0.04	16.3	21.8	0.45	5.4	6.7	0.63
Refusal	13.2	17.4	0.13	8.7	14.1	0.05	46.5	36.4	0.22	8.3	9.5	0.68
Other non-response	1.9	2.5	0.59	1.6	1.9	0.68	4.7	5.5	0.84	1.2	1.9	0.56
<i>N</i>	364	728		321	618		43	110		168	315	

Notes: F2F = face-to-face; MM = mixed modes; HH = household; *P* = *P*-values from Chi² tests adjusted for sample design.

The individual-level response rates for continuing and refreshment samples in IP5 are shown in Table 6 below. The mixed-mode design resulted in a lower proportion of individual full interviews (58.5% compared to 64.7% in the F2F group). This difference was only statistically significant for the original sample living in households who had responded at IP4. For the other two sample types, there was a higher proportion of full interviews in the mixed-mode design, but this difference was not significant. The mixed-mode design did result in fewer proxy interviews than the F2F design, possibly this was because it was not possible to do a proxy interview on-line. Again individuals in the original sample who were in households that had responded at IP4 were more likely to refuse in the mixed-mode design (14.2%) compared to the F2F design (8.5%).

Table 6: Individual Response Rates (including eligible adults in non-respondent households)

	Original sample									Refreshment sample		
	Total			IP4 responding			IP4 non-responding					
	F2F	MM	<i>P</i>	F2F	MM	<i>P</i>	F2F	MM	<i>P</i>	F2F	MM	<i>P</i>
Full interview	64.7	58.5	0.05	82.8	75.8	0.04	18.5	24.1	0.16	69.8	71.7	0.64
Proxy interview	7.3	4.4	0.00	2.6	2.0	0.50	20.2	9.4	0.00	5.5	1.9	0.00
Partial interview	0.1	1.0	0.02	0.0	1.2	0.01	0.0	0.5	0.34	0.0	1.3	0.03
Full, proxy or partial	72.1	63.9	0.01	85.5	79.0	0.06	38.8	34.0	0.30	75.3	74.9	0.93
Non-contact	7.1	5.8	0.40	3.4	2.9	0.65	18.0	13.1	0.20	5.7	7.1	0.60
Refusal	17.2	22.9	0.06	8.5	14.2	0.07	41.0	43.6	0.58	12.9	11.1	0.59
Other non-response	3.6	7.4	0.01	2.6	3.9	0.38	2.2	9.4	0.00	6.0	6.9	0.63
<i>N</i>	703	1439		495	934		178	406		348	594	

4 Experimentation in IP5

In addition to the mixed-mode design experiment, there were a number of other experiments carried on IP5 covering both fieldwork procedures and measurement in the questionnaire. There were some new experiments and some which were the longitudinal continuation of experiments carried at previous waves of the IP. This section outlines the experiments carried at IP5; briefly explaining the reasons for carrying them, describing the design of the experiment and giving an indication as to the initial results from early analysis of the data. The analyses in this working paper were based on a preliminary data-set which contained all cases but did not have weights or derived variables. The authors of each sub-section below are given in the heading.

a. **Unconditional Respondent Incentives (Peter Lynn)**

At IP5, all sample members received an unconditional incentive, enclosed with the advance letter that was mailed shortly before fieldwork began. The incentive was in the form of a High Street gift voucher, but the value of the voucher was manipulated experimentally. Members of the original sample were allocated to receive either £5 or £10, based on their randomised allocation to incentive levels at previous waves. All those who had received either £5 or £10 at IP4 again received the same amount, while those who at IP4 had received an initial £5 with a promise of an additional £5 if all household members participated were randomly allocated to receive either £5 or £10 at IP5. Members of the refreshment sample, for whom this was only their second wave, had been randomly allocated to receive £10, £20 or £30 at IP4. Each received the same value incentive again at IP5.

These experimental manipulations form part of an ongoing longitudinal experiment, the results of which will be reported separately. Here we summarise the cross-sectional outcomes at IP5.

Original Sample

At the household level, the only effect that approached even marginal significance was that in the mixed mode sample the refusal rate was perhaps slightly lower with the higher level of incentive (refusal rate 19.3% with £5 incentive, 14.8% with £10, $P = 0.09$; Table)

At the individual level, the incentive had a significant effect in the context of the mixed mode protocol: refusal rate was 19.7% with the £10 incentive, compared to 25.7% with the £5 incentive ($P = 0.05$). This led to a higher overall individual response rate, though the difference was not significant (67.6% with the £10 incentive, compared to 62.2% with the £5 incentive ($P = 0.11$)).

Amongst households that had participated at IP4, the proportion of households who participated fully by web (household grid, household questionnaire and all adult questionnaires completed by web) was higher with the £10 incentive than with the £5 incentive, but not significantly so. Amongst IP4 non-respondent households, the direction of the effect was reversed but was again not significant.

Refreshment Sample

For the refreshment sample in the face-to-face protocol, higher incentive levels were generally associated with higher response rates and lower refusal rates, at both household and individual level, though the small sample sizes mean that none of the differences are significant (

Table). With the mixed mode protocol, for which sample sizes were larger, large and highly significant effects of the incentive level were found. The proportion of households that completed all instruments was fully twenty-five percentage points higher with a £30 incentive, compared to a £10 incentive (77.7% vs. 52.9%, $P = 0.003$), mainly accounted for by a dramatic difference in refusal rates (0.9% vs. 18.4%, $P = 0.000$). Similar large differences were found at the individual level.

Table 7: Household and individual response rates by incentive level, original sample

	Face-to-face			Mixed Modes		
	£5	£10	<i>P</i>	£5	£10	<i>P</i>
Household response rate	79.5	80.4	0.83	74.5	78.1	0.23
Complete Households	58.0	60.8	0.58	50.5	53.4	0.47
Partial Households	21.5	19.6	0.69	24.0	24.8	0.84
Non-contact	6.2	3.3	0.22	3.6	5.5	0.19
Refusal	12.8	13.7	0.81	19.3	14.8	0.09
Other non-response	1.5	2.6	0.40	2.6	1.6	0.43
<i>N</i>	195	153		388	311	
Individual response rate	73.9	71.7	0.62	62.2	67.6	0.11
Full interview	65.6	65.3	0.94	58.2	60.7	0.48
Proxy	8.3	6.1	0.31	3.5	5.3	0.11
Partial	0.0	0.3	0.29	0.5	1.6	0.05
Non-contact	6.5	5.8	0.77	4.2	6.1	0.21
Refusal	16.4	18.7	0.58	25.7	19.7	0.05
Other non-response	3.2	3.9	0.66	7.8	6.6	0.49
<i>N</i>	372	311		754	639	
Household fully completing by web				19.5	24.7	0.21
<i>N</i>				330	263	

Note: Analysis restricted to IP4 responding households

Table 8: Household and individual response rates by incentive level, refreshment sample

	Face-to-face				Mixed Modes			
	£10	£20	£30	<i>P</i>	£10	£20	£30	<i>P</i>
Household response rate	87.5	87.0	90.3	0.79	74.7	81.9	92.0	0.006
Complete Households	57.5	61.1	64.5	0.78	52.9	68.6	77.7	0.003
Partial Households	30.0	25.9	25.8	0.88	21.8	13.3	14.3	0.29
Non-contact	0.0	3.7	3.2	0.48	5.8	3.8	5.4	0.78
Refusal	12.5	9.3	3.2	0.16	18.4	12.4	0.9	0.000
Other non-response	0.0	0.0	3.2	0.23	1.2	1.9	1.8	0.91
<i>N</i>	40	54	62		87	105	112	
Individual response rate	72.9	76.0	79.4	0.70	64.3	75.0	84.5	0.002
Full interview	67.1	68.6	75.4	0.49	61.2	73.5	79.7	0.007
Proxy	5.9	7.4	4.0	0.45	2.6	0.0	3.1	0.13
Partial	0.0	0.0	0.0	-	0.6	1.5	1.8	0.63
Non-contact	5.9	2.5	5.6	0.58	7.6	5.6	6.2	0.86
Refusal	16.5	17.4	6.4	0.13	21.0	13.3	2.2	0.000
Other non-response	4.7	4.1	8.7	0.37	7.0	6.1	7.1	0.94
<i>N</i>	85	121	126		157	196	226	

b. Subjective Expectations about the Returns to Schooling and the Decision to go to University (Adeline Delavande and Basit Zafar)

Higher education is a priority in UK government policy, but access to university is still unequal, with well-off families being strongly over-represented among university students (Blanden, J., and S. Machin, 2004). Several explanations could rationalize why many young individuals from lower socioeconomic background do not go to university. One possibility is that they expect low returns to a university degree. Another alternative is that they face high attendance costs or binding credit constraints. Without data on perceptions and expectations of individuals, it is not possible to distinguish between these explanations (e.g., Manski, 2004). Yet, any policy aimed at promoting schooling can only be successful if the main reasons for not enrolling in university are understood.

In IP5, we collected expectations from youth aged 16 to 21 currently not at university, and from parents of children aged 10 to 21 about their expectations for their eldest child. In particular, we asked about: (1) intention to apply to university (the percent chance on a scale from 0% to 100% of gaining the qualifications to go to university; the percent chance of applying to university, and; the percent chance of applying to university if all costs were paid out of scholarship), (2) expected success at university (the percent chance of graduating conditional on applying), (3) the financial returns to a university degree (expected earnings at age 30 conditional on going to university in first field of study and conditional on not going to university, and percent chance of being employed at age 30 conditional on going to university in first field of study and conditional on not going to university), (4) the expected monetary costs of going to university (expected tuition and expected loan), and (5) the expected non-monetary cost of going to university (expected hours of study).

Table 9 presents the average expectations for parents and children for the whole sample, and for respondents coming from high or low income households (defined as those above and below the household median income in the sample, respectively) along with response rates. The differences in aspirations by socioeconomic status are clear in the very first question, which elicits beliefs about qualifying for university: while 78 percent of parents belonging to high-income households believe their child would gain the required qualifications for University, only 67 percent of lower-income households believe so (difference statistically significant at the 5% level). A similar difference is seen in the chance of applying. Interestingly, providing scholarship that covers all costs essentially closes the gap in intention to apply by socioeconomic status: the average chance of applying with scholarship is very similar for low and high income respondents, for both parents and children. This suggests that financial constraints play a role in poorer students' intention to apply.

The table also shows that respondents perceive a positive payoff to a university degree versus no university degree: the expected average university premium (expected age 30 earnings conditional on a university degree - expected age 30 earnings conditional on no university education) is £13,300 for parents and £12,100 for children, and the premium in terms of probability of employment at age 30 is 6.4 percentage point for parents and 12.5 percentage points for children. Notably, respondents from different income groups perceive different returns. For example, parents from high-income households expect a larger average university premium: £14,700 versus £11,800 for low-income household parents.

Table 9: Child and parental expectations

Variable name	Parental Response				Child Response			
	Full Sample	High ^a Income	Low Income	Response Rate (%)	Full Sample	High ^a Income	Low Income	Response Rate (%)
Chance of qualifying for university	72.1	78.0**	66.8	96.6	71.3	77.1	67.0	98.4
Chance of applying to university	71.0	74.1	67.4	97.4	69.2	71.0	66.6	98.2
Chance of applying to university Scholarship	79.1	80.3	77.2	96.2	72.9	70.0	74.2	100.0
Chance of completing university degree	89.2	92.1	86.5	96.6	93.0	91.4	94.3	100.0
Expected age 30 earnings university degree (on £1000's)	37.5	40.6*	34.3	83.6	34.6	36.3	33.9	94.6
Expected age 30 earnings no university (on £1000's)	24.2	25.9***	22.5	81.9	22.5	26.1***	19.2	90.3
Hours expected to study	19.8	20.4	19.3	75.3	22.0	24.5	19.4	89.4
Chance of being employed university degree	82.0	83.2	81.2	91.8	85.0	84.6	86.8	98.2
Chance of being employed no university	75.6	77.6	72.8	89.8	72.5	80.4	66.5	95.2
Expected Annual Tuition (in £1000's)	7.7	7.6	7.7	63.4	8.4	9.3	7.2	71.4
Expected loans if attend university (all years, in £1000's)	17.0	18.0	16.6	58.3	16.0	19.3*	11.8	60.7
Maximum observations	256	128	111		61	30	27	

Table reports pairwise Wilcoxon tests for equality of means for high versus low income (stars shown on the high income column).

***, **, * denote significance at the 1, 5, and 10% levels, respectively.

^a Low (high) income is a household with below (above) median household income.

c. Persuasion Messages in Advance Letter (Cong Ye)

In longitudinal surveys, because the survey organization has made successful contacts with the sample members in previous rounds, the challenge to locate and make contact with the sample members is usually not a major problem. Most nonresponse is often due to refusal.

The effort of obtaining cooperation from sample persons is a persuasion effort; therefore, the persuasion strategies are important in gaining cooperation from sample members. This experiment was designed to test the effects of persuasion messages on response rates and measures of helping attitude and conformity. According to the self-perception theory (Bem, 1972), people may infer from their compliance with the previous request that they are the kind of persons who do such things. If we label respondents as “helpful persons,” this may encourage this inferential process (Strenta and Dejong, 1981). Moreover, survey researchers can provide norm information (e.g., what percent of sample members participated) to influence the sample member’s decision to take part, because many people have a tendency to act in a same manner as they believe most people do (Gorassini and Olsen, 1995). In the advance letters, one-quarter of sample members received an additional sentence “your responses in previous survey show that you are a helpful person”, one-quarter received an additional sentence “almost everyone like you responded in the last wave of the survey”, one-quarter received both sentences, and the final quarter acted as the control group, receiving neither of them.

Table 10 shows that there is variation in response among the four experimental groups. However, this is mainly because the “Norm” group (almost everyone like you responded in the last wave of the survey) has a lower response rate. A backfire effect is observed for this priming. The “helpful” priming had little impact on response rate, although it seems to mitigate the backfire effect from the “norm” priming. This backfire effect is more obvious if we compare the “Norm” group and the “Both” group to the other two groups, as shown in Table 11.

Table 10: Response Rate by Experimental Group

	Helpful	Norm	Both	Neither	Total
Response Rate	80.4%	73.4%	77.4%	79.5%	77.7%
<i>n</i>	383	383	381	380	1,527

$\chi^2(3): 6.46, p < .10.$

Table 11: Response Rate by “Norm” Priming or Not

	Norm/Both	Other	Total
Response Rate	75.4%	80.0%	77.7%
<i>n</i>	764	763	1,527

$\chi^2(1): 4.57, p < .05.$

However, controlling for response status in the previous wave, Table 12 shows no significant effects. The overall backfire effect is mainly due to the fact that previous wave non-respondents who responded in a much lower rate in the “Norm” priming group.

Table 12: Response Rate by “Norm” Priming or Not by Response Status in Previous Wave

	Prev. Respondents			Prev. Nonrespondents		
	Norm	Other	Total	Norm	Other	Total
Response Rate	81.2%	83.7%	82.4%	29.1%	39.1%	33.6%
<i>n</i>	679	698	1,377	85	64	149
	$\chi^2(1): 1.51, p=0.22.$			$\chi^2(1): 1.53, p=0.22.$		

No obvious patterns are observed for persuasion effects on measures of helping attitude and conformity.

d. Satisfaction and the implicit comparison group (Laura Fumagalli)

The study of happiness and satisfaction has led to a few puzzles. For example, contrary to what standard economic theory predicts, increased material prosperity and improved life standards do not seem to translate into higher levels of self-reported life satisfaction or happiness (see: Easterlin, 1974, 1995; Blanchflower and Oswald, 2004).

In particular, this is the case for women, whose well being has not risen, in spite of the consequences of anti-female discrimination policy (Blanchflower and Oswald, 2004), and the enormous gains they experienced as a group. Steveson and Wolfers (2009) analyse a large sample of countries (including most western European countries and the USA) and find that women’s happiness has generally failed to keep the pace with women’s emancipation, as in most cases women’s self reported happiness has decreased relative to men’s, and, in a few cases, also in absolute terms. Two main explanations have been proposed for why women seem to get more unhappy relative to men. Among the most convincing explanations, it has been pointed out that the puzzle can be explained by a change in the implicit comparison group, if women started raising their expectations up to the level of men’s they became increasingly frustrated when their achievements did not meet their expectations.

A similar argument is also used by Clark (1997) who observes that less educated women seem to be more optimistic than more educated women in reporting their level of job satisfaction, and concludes that this must happen because education increases expectations. The importance of the implicit comparison group in answering questions about (job) satisfaction is also stressed by Clark and Oswald (1996), who find that job satisfaction decreases with education. They suggest that an explanation can be that the recession of the early nineties disproportionately hit the middle class and the “low satisfaction of the higher educated might then just result from their sharp drop in income relative to those with lower qualification”. As an alternative explanation, they suggest that the result is due to education rising expectations, thus decreasing satisfaction.

Design

The experiment tries to study whether the level of reported satisfaction depends on the comparison group to which people implicitly refer when they answer questions on satisfaction. In particular, we asks the respondents to answer a set of questions on satisfaction in different aspects of life (namely: job, health, household income, leisure and life overall) by forcing them to refer to a particular comparison group. We use an interpenetrated design with two treatments defining the comparison group on the basis of 1) gender, 2) education. This results in four different treatment groups (see table 13), which are randomly allocated to sample members.

Table 13: treatments.

	Education=1	Education=0
Gender=1	How dissatisfied or satisfied are you with aspect of life if you compare yourself with other men/women with your level of education?	How dissatisfied or satisfied are you with aspect of life if you compare yourself with other men/women?
Gender=0	How dissatisfied or satisfied are you with aspect of life if you compare yourself with other people with your level of education??	How dissatisfied or satisfied are you with aspect of life?

For all these questions, the respondents can choose one of the following answers: 1. “Completely dissatisfied”, 2. “Mostly dissatisfied”, 3. “Somewhat dissatisfied”, 4. “Neither satisfied nor dissatisfied”, 5. “Somewhat satisfied”, 6. “Mostly satisfied”, and 7. “Completely satisfied”.

Hypotheses and empirical specification

If some women compare themselves to men, and this makes them report a lower level of satisfaction, we would expect that forcing them to compare themselves to other women will increase such levels. This would be more pronounced in cases where women are still likely to have worse outcomes than men (namely for job satisfaction, and income).⁷ Following Clark (1997), we would also expect that such comparison raises the levels of job satisfaction for high educated women. Moreover, if it is true that the amount of leisure women can enjoy has reduced if compared to men’s (see see Krueger, 2007) then, forcing women to compare themselves to other women should also increase satisfaction on leisure. The effect on health is not clear, and it depends on people’s beliefs on gender health gradient. Finally, based on Clark and Oswald’s hypothesis about the recession hitting disproportionately the higher educated, we expect that forcing respondents to compare themselves to people with the same level of education would lead to an increase in (job) satisfaction for highly educated people. However, if their hypothesis on education raising expectations is correct, then forcing respondents to compare themselves to people with the same level of education would lead to a decrease in (job) satisfaction for highly educated people.

Although generally the literature has focused on the share of respondents who answer they are “completely/very satisfied” or “completely/very” or “mostly satisfied”, our sample size does not

⁷ Since we just have household income, we will focus on single respondents only

permit to do this, due to small cell sizes. As a consequence, we focus on two variables: one indicating whether the respondent is “Satisfied” (answers 1 to 3) and the other indicating whether the respondent is “Dissatisfied” (answers 1 to 3). We run each model by gender and by level of education (university, post compulsory and no or compulsory education)

Table 14: Effect of the comparison by gender, by gender (selected coefficients)

	Males	Females	Males	Females
	Satisfied		Dissatisfied	
Job				
Treatment:	0.3922	0.5557**	-0.0413	-0.1591
Gender	(0.256)	(0.246)	(0.396)	(0.371)
N	529	586	529	586
Health				
Treatment:	-0.1212	0.1204	0.0608	-0.4872**
Gender	(0.212)	(0.195)	(0.231)	(0.211)
N	779	949	779	949
Income				
Treatment:	-0.3204	0.5740**	0.0715	-0.7397**
Gender	(0.317)	(0.280)	(0.333)	(0.311)
N	321	433	321	433
Leisure				
Treatment:	0.1208	0.2958	-0.3241	-0.3928*
Gender	(0.211)	(0.192)	(0.241)	(0.218)
N	775	951	775	951
Overall				
Treatment:	-0.1351	-0.3232	0.1577	-0.1474
Gender	(0.226)	(0.207)	(0.275)	(0.257)
N	775	952	775	952

Standard errors in parentheses. *= $p < 0.10$, **= $p < 0.05$, ***= $p < 0.01$.

Table 14 gives some indication that the implicit comparison group plays a role in the reported level of satisfaction. These results are particularly significant in the case of income, in spite of the smaller sample size due to the exclusion of the respondents in couples. In fact, women who are asked to compare themselves to other women are significantly more likely that women in the control group to answer that are satisfied about their income and significantly less likely to answer that they are dissatisfied. Moreover, when comparing themselves to other women, women are more likely to be satisfied about their job, and less dissatisfied about their health and leisure. Perhaps surprisingly, no effects are found on satisfaction about life overall.

Table 15 is meant to test the hypothesis, based on Clark (1997), that the effect of forcing women to compare themselves with other women should increase with education. It shows that the effect is actually driven by women with post-compulsory non-university education, with gives only some

weak support to the hypothesis. Moreover, a similar (though smaller) effect is also found for males, which is not in line with the theories of the implicit comparison group.

Table 15: Effect of the comparison by gender, by gender and education (selected coefficients)

	High Education		Middle Education		Low Education	
	Males	Females	Males	Females	Males	Females
Satisfied						
Treatment:	0.0870	0.0412	0.7129*	1.3204***	0.0382	0.4472
Gender	(0.461)	(0.435)	(0.375)	(0.407)	(0.677)	(0.579)
Dissatisfied						
Treatment:	-0.0488	-0.6190	-0.2318	0.0852	0.8755	-0.8224
Gender	(0.691)	(0.623)	(0.538)	(0.541)	(1.274)	(1.189)
N	183	199	240	234	78	121

Standard errors in parentheses. *=p<0.10, **=p<0.05, ***=p<0.01.

Table 16 focuses on the comparison with people with the same level of education. Again, forcing respondents to compare themselves to people with the same level of education increases the probability of reporting a higher level of job satisfaction for respondents with post-compulsory, non-university education.

Table 16: Effect of the comparison by education, by education (selected coefficients)

	High	Middle	Low	High	Middle	Low
		Satisfied			Dissatisfied	
Job						
Treatment:	-0.0500	0.8066***	-0.053	-0.0899	-0.2046	-0.1987
Education	(0.295)	(0.271)	(0.413)	(0.419)	(0.393)	(0.792)
N	382	474	199	382	474	199
Health						
Treatment:	0.3407	-0.2183	-0.1823	-0.6240**	-0.2501	-0.4210
Education	(0.272)	(0.225)	(0.301)	(0.310)	(0.251)	(0.327)
N	487	701	381	487	701	381
Income						
Treatment:	-0.1809	0.2816	0.4032	-0.2763	-0.5477*	-0.0096
Education	(0.344)	(0.306)	(0.390)	(0.424)	(0.328)	(0.415)
N	312	354	226	312	354	226
Leisure						
Treatment:	0.1884	0.1542	-0.1218	-0.3851	-0.1929	0.2973
Education	(0.261)	(0.220)	(0.301)	(0.305)	(0.242)	(0.323)
N	487	700	381	487	700	381
Overall						
Treatment:	0.0333	-0.5759**	-0.4132	-0.4532	0.4148	0.4129
Education	(0.294)	(0.239)	(0.320)	(0.372)	(0.297)	(0.374)
N	487	700	382	487	700	382

Standard errors in parentheses. *=p<0.10, **=p<0.05, ***=p<0.01.

This could be in line with Clark and Oswald (1996) if the recession, which also affected our sample, has disproportionately hit such people. Clark and Oswald could also explain why this effect does not extend to people with university degree, since high education is also likely increases expectations, thus counterbalancing the effect of the comparison. A similar effect for people with a middle level

of education is also found in the case of dissatisfaction for income. Finally, a decrease in the probability of being dissatisfied with health for people with a university degree and a decrease in overall life satisfaction for respondents with post-compulsory non-university education are found. These effects can hardly be explained based on theories on the implicit comparison group.

Conclusion

The experiment gives some evidence that some of the differences in reported levels of satisfaction could be given by differences in the implicit comparison group. However, the power of the experiment is not enough to suggest definite conclusions. This experiment was initially meant to be carried out for at least two waves, but it was discontinued afterwards. The results suggest that it may be worth repeating it.

e. **Wording effects of dependent interviewing questions on the amount of change observed in panel data (Annette Jäckle)**

This study was designed to investigate how best to word dependent interviewing questions. *Understanding Society* uses dependent interviewing for many questions. For some items respondents are reminded of their answer in the previous interview and asked *whether this is still the case*, for other items they are reminded and then asked *whether this has changed*. In IP3/IP4 we implemented an experiment contrasting these two versions. Preliminary analyses showed clear evidence of agreement bias: respondents were much more likely to say “Yes, this is still the case, than “No, this has not changed” – and much more likely to say “Yes, this has changed” than “No, this is not still the case” (see Burton et al, 2012).

In this follow-up study we contrasted the two question formats with a ‘balanced’ format which does not allow simple confirmation. To study potential response order effects, and whether these differ in CAPI and web, we varied the order of response options in the new format.

Respondents were randomly allocated to one of four question versions. In each version respondents were first reminded of the answer they had given in the previous wave, for example: “*Last time you said that you were salaried {if ff_paytyp = 1} / you received a basic salary plus commission {if ff_paytyp = 2} / you were paid by the hour {if ff_paytyp = 3}.*” This reminder was followed by a question about whether the status had changed:

Group 1: “*Is this still the case?*” (Yes/No)

Group 2: “*Has this changed?*” (Yes/No)

Group 3: “*Is this still the case or has it changed?*” (Still/Changed)

Group 4: “*Has this changed or is it still the case?*” (Changed/Still)

This experiment was implemented on a number of items in the individual questionnaire.⁸ The independent questions asked in IP4, of which respondents were reminded in this experiment, were:

LKMOVE: *“If you could choose, would you stay here in your present home or would you prefer to move somewhere else?”*

- 1 Stay here*
- 2 Prefer to move*

EDTYPE (if full-time student) *“Are you...”*

- 1 At School*
- 2 At Sixth Form College*
- 3 At Further Education (FE) College*
- 4 At Higher Education (HE)*

PAYTYP (if employee) *“How is your pay calculated, in particular are you salaried or paid by the hour?”*

- 1 Salaried*
- 2 Basic salary plus commission*
- 3 Paid by the hour*

WORKTRAV (if employee and works somewhere other than home) *“And how do you usually get to your place of work?”*

- 1 Drive myself by car or van*
- 2 Get a lift with someone from household*
- 3 Get a lift with someone outside the household*
- 4 Motorcycle/moped/scooter*
- 5 Taxi/minicab*
- 6 Bus/coach*
- 7 Train*
- 8 Underground/Metro/Tram/Light railway*
- 9 Cycle*
- 10 Walk*

JSPART (if self-employed) *“Are you working on your own account or are you in partnership with someone else?”*

- 1 Own account (sole owner)*
- 2 In partnership*

JSWORKTRAV (if self-employed and works somewhere other than home) *“And how do you usually get to your place of work?”*

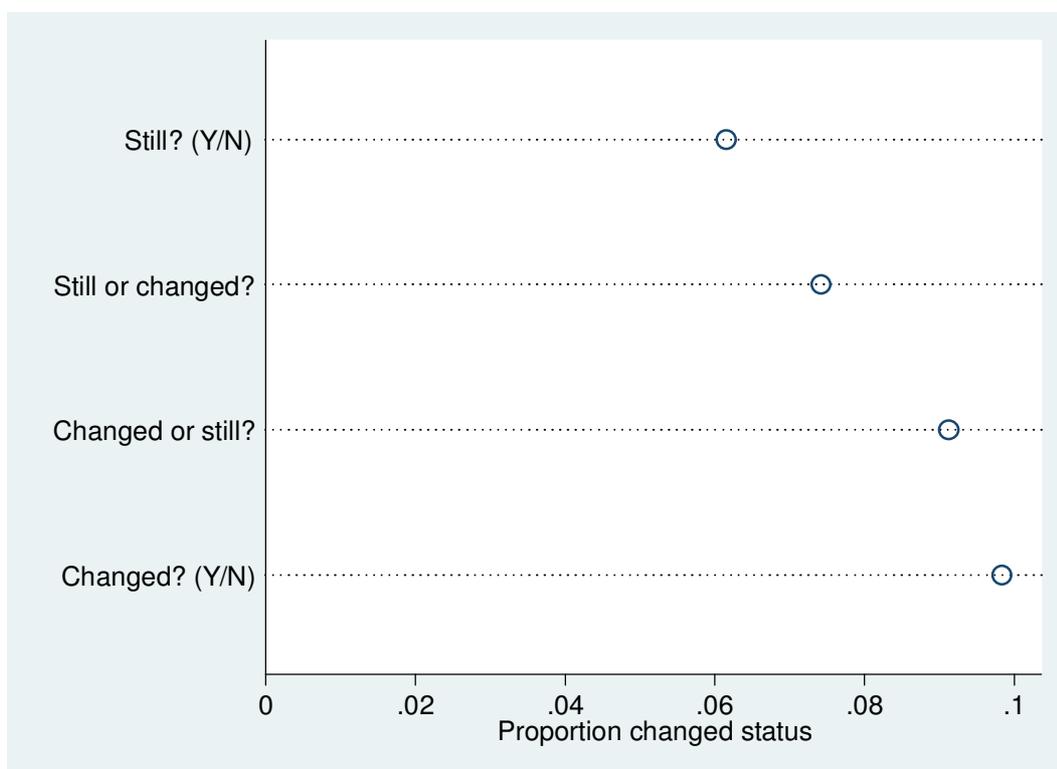
[Response options as for WORKTRAV]

⁸ The experiment was implemented on some additional items for which the implementation was corrupted. See the Innovation Panel User Guide for documentation of the errors. This included items in the household questionnaire (number of rooms, housing owned or rented, expenditure on mortgage, expenditure on rent), and some further items in the individual questionnaire (whether job permanent, number of employees at workplace).

The experiment was implemented on some additional items in the individual questionnaire for which dependent interviewing had already been used in IP4: general health, occupation, industry, whether employee or self-employed, and number of hours worked. These items are excluded from the analyses presented here, as the preload in the experiment was potentially already affected by the use of dependent interviewing in IP4. Although not shown, the results including these items are similar to those presented here.

Figure 1 plots the rates of status changes reported across all experimental items, depending on the wording of the dependent interviewing question. Respondents were least likely to report a change if asked “*Is that still the case?*” (6.2%), and more likely to report a change if asked “*Has that changed?*” (9.8%, $P=0.007$), or “*Has that changed or is it still the same?*” (9.1%, $P=0.023$). The graph suggests that respondents were more likely to pick the first response option in the balanced questions, i.e. to report a change if asked “*Has that changed or is it still the same?*” and to report no change if asked “*Is that still the same, or has that changed?*” This differences was however not significant, and there did not appear to be any difference in this potential response order effect between respondents who completed the survey by web and those who completed in face-to-face interviews.

Figure 1: Proportion of respondents reporting a status change, by dependent interviewing question wording



Concerns about data confidentiality with dependent interviewing

This study was designed to investigate whether respondents react differently to the use of dependent interviewing in a web survey than in face-to-face surveys. A previous study in the context of the

British Household Panel Survey had found that face-to-face respondents had little concern about their responses from previous waves being preloaded into the interview. In fact they expected interviewers to have access to their previous data (Sala et al, 2011; see also Pascale and Mayer, 2004). In a web survey it is however possible that respondents may be more concerned about the confidentiality of their responses if they are preloaded into the survey.

At IP5, respondents in the refreshment sample were interviewed for the second time. For these respondents it was the first time they were asked questions using dependent interviewing. To test their reactions all refreshment sample members were asked two additional questions at the end of the self-completion section:

DIREAC1: *“In this interview there were some questions where you were reminded of answers that you had given in your previous interview. How concerned are you about the confidentiality of your answers, knowing that we may use them in a later interview?”*

- 1 Very concerned*
- 2 Somewhat concerned*
- 3 Not concerned*

DIREAC2: *“Did reminding you of your previous answer affect how willing you are to take part in future interviews for this survey? Did the reminders make you...”*

- 1 More willing*
- 2 Less willing*
- 3 Made no difference*

Respondents who completed the survey by web expressed more concern about the confidentiality of their data.⁹ As Figure 2 shows, 32.8% of web respondents expressed either strong or some concern, compared to 22.5% of face-to-face respondents (P=0.010).

Further, Figure 3 shows that face-to-face respondents were more likely to say that the use of dependent interviewing made them more willing to participate in future waves of the survey (12.2% face-to-face vs. 9.7% web), while web respondents were more likely to say it made them less willing (2.1% face-to-face, vs. 6.2% web, P=0.027).

⁹ Note that this analysis was done by the mode respondents completed the survey in, and does not account for the self-selection into mode for the mixed mode sample, where sample members who did not participate by web were then followed up face-to-face.

Figure 2: Concern about data confidentiality, by mode of interview

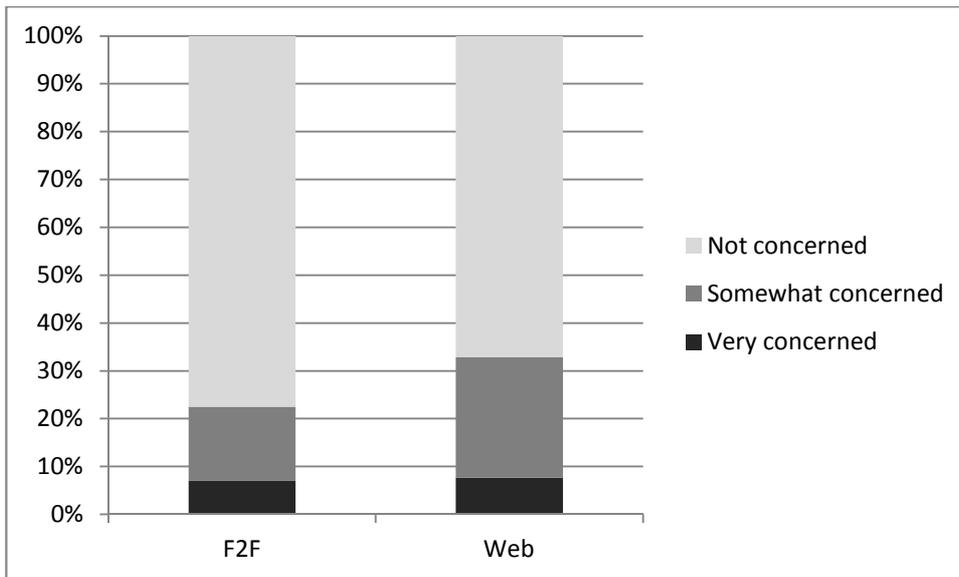
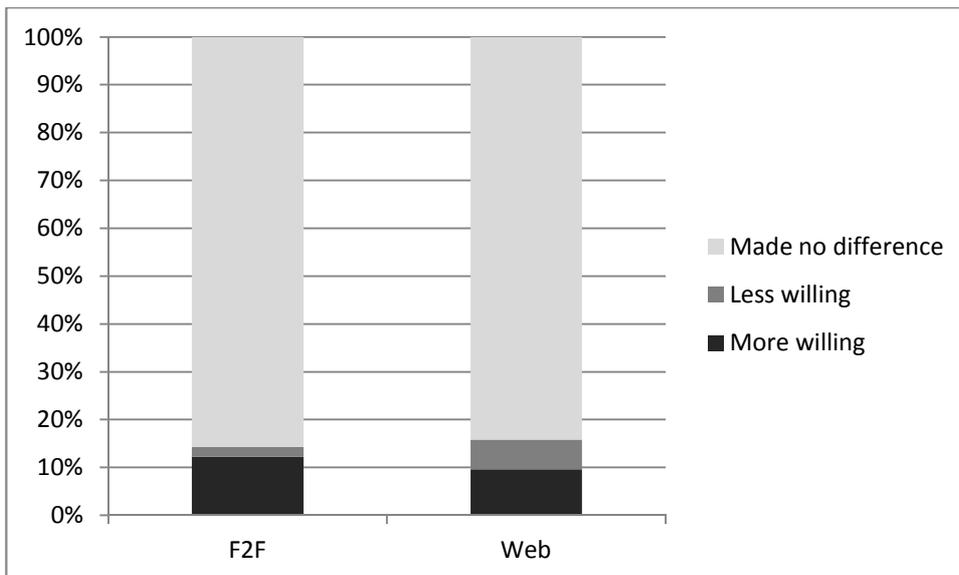


Figure 3: Effect of using dependent interviewing on willingness to participate in future interviews, by mode of interview



f. Tailoring mode of data collection in longitudinal studies (Olena Kaminska and Peter Lynn)

An optimal balance between survey costs and participation rates might be achievable if we knew the mode(s) in which each sample member was most likely to participate and would then administer the survey in the optimal mode(s) for each sample member. In other words we would use a cheaper mode for those sample members predicted to be just as likely to respond in that mode and we would use a mode with a higher predicted response propensity for sample members predicted to be less

likely to respond in the standard mode. This study assesses the feasibility and effectiveness of such mode tailoring.

At its simplest, this could involve asking a single question about mode preference, as some previous studies have done. The strength of this approach is its simplicity. The limitations of the approach include an inability to identify the ranking of modes other than the preferred one and an inability to identify the magnitude of differences in preferences between modes. A slight refinement would be to ask separate questions about the likelihood of responding in each mode, to provide a simple estimate of response propensity in each mode. We asked variants of both forms of the question, as in principle they provide complementary information. First, respondents were asked to pick their most and least preferred modes among four modes (face-to-face, telephone, postal self-completion and web). In addition, using a scale from 0 to 10, respondents were asked to rate their likelihood of responding in the future if contacted in each of three modes: telephone, postal and web. We randomized the order of asking mode preferences before and after participation likelihood and reported results in Budd et al. (2012).

The questions on mode preference were administered in face-to-face mode at wave 4 of the Innovation Panel. At wave 5 the sample was randomly assigned to either face-to-face mode or a mixed mode (MM) protocol. The MM protocol invited panel members to participate via web mode first, and followed with face-to-face mode for those who didn't respond to web mode within two weeks.

Results: predictive power of mode preference on participation in different modes

The most interesting question is whether self-reported mode preference in a previous wave can predict response likelihood in the following wave. We analyse three types of response rate: response rate for a face-to-face condition, for MM condition and for web part only in the MM condition (Table 17).

Table 17: Response rate in face-to-face, mixed-mode (MM) and web part of mixed mode conditions depending on mode preferences

	RR f2f	RR MM	RR web	RR difference	N	percent
Web participation likelihood						
Definitely would not do 0	87.1	77.3	15.3	9.9	672	28.4
1	88.0	81.3	25.0	6.8	57	2.4
2	76.2	80.0	36.0	-3.8	71	3.0
3	84.0	80.0	27.3	4.0	80	3.4
4	87.5	83.0	51.1	4.5	55	2.3
5	82.0	72.5	44.0	9.5	159	6.7
6	74.1	82.8	57.8	-8.7	91	3.9
7	82.2	76.8	51.2	5.4	127	5.4
8	74.7	76.0	51.2	-1.4	196	8.3
9	81.5	78.9	61.5	2.7	169	7.2
Definitely would do 10	83.5	77.5	52.0	6.1	466	19.7
Most preferred mode						
Face-to-face	87.0	80.0	32.6	7.0	1286	54.4
Telephone	90.0	72.2	33.3	17.8	28	1.2
Self-completion	75.8	74.5	36.2	1.3	324	13.7
Web	78.0	73.8	55.8	4.2	480	20.3
no preference (vol)	83.3	63.2	47.4	20.2	31	1.3
Least preferred Mode						
Face-to-face	66.7	64.4	33.3	2.2	60	2.5
Telephone	81.8	77.9	44.7	4.0	1405	59.4
Self-completion	80.0	74.1	38.1	5.9	219	9.3
Web	89.8	81.5	20.3	8.3	438	18.5
no preference (vol)	90.0	60.0	33.3	30.0	25	1.1

As can be seen, mode preference is a good predictor of participation in the web part of a mixed mode data collection. Among respondents who rate their likelihood to participate via web above 6 on a 0 to 10 scale, the response rate in the web part of the MM protocol is over 50%, while those who say that they definitely will not participate have a response rate of 15.3%. Similarly, respondents whose most preferred mode is web have a 56% chance to respond via web in the following wave, compared to 33% among those who prefer face-to-face or telephone modes. And respondents who prefer web mode least have 20% of chance to participate by web, which is considerably lower than the response rate among respondents reporting other modes as least preferred.

Nevertheless, mode preference seems to be a less useful predictor of the overall mixed mode response rate in which web mode is followed by face-to-face mode. For example, while response rates for MM and face-to-face protocols are similar for those who rate their likelihood to participate via web as 8 or 9 (differences of less than 3 percentage points), the difference is 6.1% for those

whose reported likelihood as 10. The percentage point difference in response rates between face-to-face and MM conditions ranges from 9 in favour of the MM condition to 30 in favour of the face-to-face condition.

The decision about whom to assign to MM or face-to-face condition should depend on two factors: the cost and the difference in response rate. It is hard to compare costs between conditions as the web often has high starting cost, but very little cost per each additional interview, unlike face-to-face where each interview requires interviewer time and travel expenses. The cost function for face-to-face mode is further complicated by the need to interview all individuals within a household, often resulting in either multiple visits or incomplete households. These considerations should be taken into account when tailoring mode assignment to different groups.

Note, that all respondents indicated their likelihood of participation in web, and their least and most preferred modes. It is the combination of these measures and possibly other variables that may be found to be most useful in tailoring modes of data collection. Future research should explore these possibilities.

g. Weekday of Advance Mailing and Conditional Incentives Experiment (Carl Cullinane and Gerry Nicolaas)

This experiment examined two ways of maximising take-up of the IP5 web instrument, through the means of timed advance letters and emails, and a conditional incentive offered to households if all members of the household completed the survey online. These treatments were combined into a four-way experiment.

In order to explore whether the timing of the advance letters and emails had an effect on response, half of the IP5 web sample (n=1077) were sent advance mailings (by post, and email where possible) timed to arrive on a Monday, and the other half on a Friday.

The conditional incentive experiment aimed to explore whether a conditional incentive, offered in addition to existing unconditional incentives, would increase take up of the web survey. To this end, half of the web sample was offered a £10 bonus per adult household member if the entire household completed online, with the other half of the sample receiving just upfront incentives. This allocation was crossed with the advance mailing treatment groups. The conditional incentive experiment is particularly important, as encouraging full households to complete the survey online eliminates the necessity of an interviewer calling and thus maximises the cost savings of a mixed mode approach.

The base for these figures includes all households allocated to the Web sample, composing members of both the original IP sample and the refreshment sample.

Table 18: Household web response rate, full web sample.

	Monday, bonus	Monday, no bonus	Friday, bonus	Friday, no bonus	Total
Partial and unproductive household	77.5%	79.6%	72.9%	79.6%	77.4%
	214	211	194	215	834
Full productive household	22.5%	20.4%	27.1%	20.4%	22.6%
	62	54	72	55	243
Total	276	265	266	270	1077

Table 18 shows the results for the combined advance mailing and conditional experiments. The experimental manipulations seemed to have a limited effect on response, with Friday mailings combined with the incentive bonus the only category to display a marked contrast from the others, suggesting a potential interaction effect on response.

Day of Advance Mailing

For households who received advance mailings on a Monday, the overall proportion of full households completing by web was 21.4%, while for those receiving on a Friday, response was 23.7%. This difference however was not statistically significant.

All incomplete cases and non-responses were then passed to face-to-face interviewers. While the pattern also held for final household response (Monday: 51.9%, Friday: 54.7%), this also fell some way short of statistical significance.

Conditional Incentive Experiment

A greater degree of difference emerged from the conditional incentive experiment however. Full household web response from households offered no bonus was 20.4%, while households who were offered the £10 bonus had a response rate of 24.7%. This effect fell just below the threshold of significance at the 5% level, but was significant at the 10% level, offering tentative evidence that the web bonus was effective at increasing full household web response.

However, when the sample is restricted to just the long-standing members of the panel and excludes the refreshment sample (n=756), the effect of day of mailing remains the same, but the effect of the web bonus is more pronounced. 21.1% of households offered the bonus responded fully by web compared to only 14.5% of non-bonus households ($p < .05$). This suggests that the bonus had a markedly greater effect among established, long-standing members of the panel than the newer additions.

Nonetheless, this effect on response had disappeared by the time face-to-face fieldwork among web non-respondents had been completed. Final household response rates for both those who were offered the web bonus and those who were not were almost the same, at 53% and 53.6% respectively. This pattern holds for both the original and refreshment samples.

So while there is some evidence that implementing a web bonus strategy could potentially cut costs by increasing full household take up of a web survey, there is no such evidence that this has any effect on final response rates.

h. Measuring Partnership Satisfaction with the Division of Housework (Katrin Auspurg, Maria Iacovou, Cheti Nicoletti)

The purpose of this experiment was to test empirically the reasons behind the gendered division of housework; in particular, to assess whether the gendered division of labour relates to differences in the utility derived by men and women from different divisions of labour within a couple. In conventional household surveys, one observes only individuals' satisfaction with their *actual* arrangements, meaning that women are rarely observed to do more paid work, or to earn more, than their male partners. Thus, it is difficult to assess whether women's greater contribution to housework arises because of gender norms, or as the result of a process of specialisation resulting from differences in skills or preferences.

There are obvious difficulties with randomised experiments: in the real world, it is not possible to randomly allocate paid work, earnings, children or housework between members of couples. However, we can ask people to imagine themselves in different scenarios, and to tell us how content they would be with different sets of arrangements.

People's satisfaction with the amount of housework they do may vary with a range of factors, including how much paid work they and their partners do; with the level of their own and their partners' earnings; with the presence and ages of children in the home; and whether the household employs paid help (for example, in the form of a cleaner). The scenarios which we put to respondents vary across all of these dimensions, following the technique of factorial survey designs.

In generating the survey questions, each of the factors mentioned above was categorised into between two and five levels, as shown in Table 19.

Table 19: Vignette dimensions and levels

Dimensions	Levels					In total
	1	2	3	4	5	
1 Amount of paid work	both full time	both half time	resp. full time, partner half time	partner full time, resp. half time		4
2 Hourly pay	partner twice as high as resp.	resp. twice as high as partner	Approx. same amount			3
3 Number and age of children	0	1 aged 6 months	1 aged 5 years	1 aged 15 years		4
4 Own share of housework	Nothing	a quarter	half	three quarters	all	5
5 Paid housework	no	one morning per week				2
Vignette universe (4x3x4x5x2)						480

A battery of questions was generated, spanning all possible combinations of these explanatory factors. A sample question reads as follows:

“Imagine that you are married or cohabiting, you and your partner both have full time jobs, and your hourly pay is approximately the same as your partner’s. You have one child aged 5 years; your partner does one quarter of the housework while you do three quarters of it, and you do not employ anybody to help with the housework.”

How satisfied would you say you are with the sharing of the housework?

Respondents were asked to reply on a seven-point scale, from 1 (completely dissatisfied” to 7 “completely satisfied”.

Each respondent received three questions selected at random from this battery.

Brief Description of Responses to Vignette Questions

Randomisation

It is important for the factorial survey method to work well that the three questions received by each respondent be random, in that (a) the questions received be uncorrelated with the characteristics of the individual; (b) the factors varying between questions be not cross-correlated; and (c) each level of each of the factors occur with approximately equal frequency.

We checked whether these conditions held in the sample of respondents, and found that they hold – the correlations with a set of four personal characteristics (age, sex, marital status, children) and three couple characteristics (actual satisfaction with housework, both partners’ hours of housework, and between-partner differences in standards of housework), are all below 0.04. The cross-correlations (b) are even lower, and there is almost perfect balance (c) between the levels of each of the factors.

Response rates

The table below presents item non-response rates for the vignette questions. Only 4% of respondents did not answer any of the three questions; the percentages responding to only one or two questions was also low, at 1.4% and 1.8% respectively.

Table 20: Valid vignette ratings for respondents participating in FS module

	<i>N</i>	<i>%</i>
0	65	4.0
1	23	1.4
2	30	1.8
3	1494	92.7
Total	1612	100.0

This means that almost 93% of respondents gave valid responses to all three vignette questions. However, this “headline” figure may mask other problems. Table 21 shows that of those who gave valid responses to all three questions, one-third gave identical responses to all three.

Table 21: Number of different ratings, only respondents with 3 valid ratings

	<i>N</i>	%
All three identical	483	32.3
Two the same, one different	657	44.0
All three different	354	23.7
Total	1612	100.0

This does not necessarily indicate a problem, since respondents may genuinely be indifferent between scenarios. However, further investigation reveals that a process of fatigue may be affecting results.

Figure 4 shows the distribution of responses across all respondents. This differs from patterns of satisfaction typically observed in questions in British household surveys, where the second-highest category is normally the mode, and where only a small minority of respondents record responses in the lowest category. The distribution of responses to the vignette questions is much flatter, although because they are responses to hypothetical scenarios, we may expect that they would show a higher variance.

Figure 4: Vignette ratings

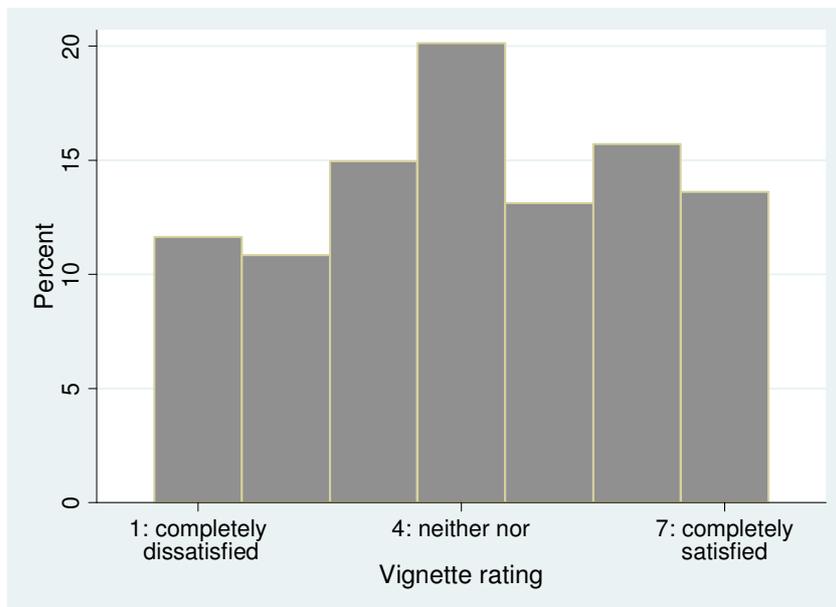
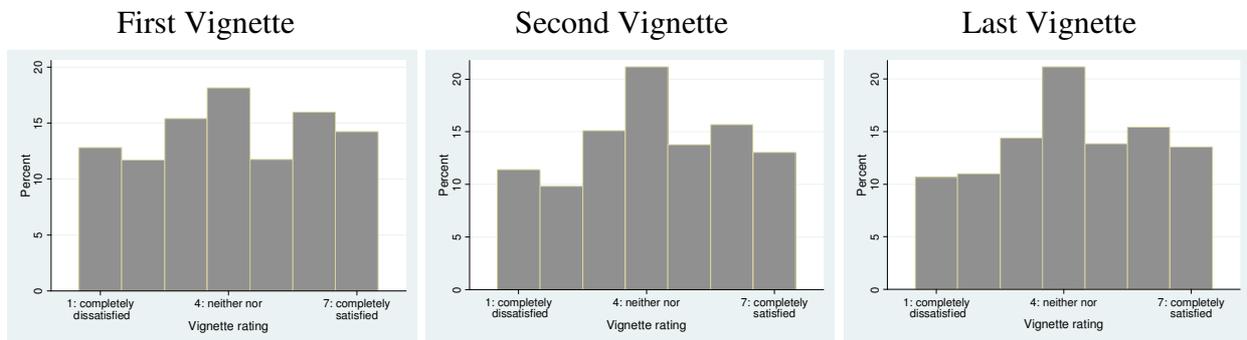


Figure 5 shows the distribution of responses, by the order in which questions were asked. There is clear evidence of clumping at the midpoint of the scale by the last question, suggesting that respondents may be giving the second, and particularly the third question, less

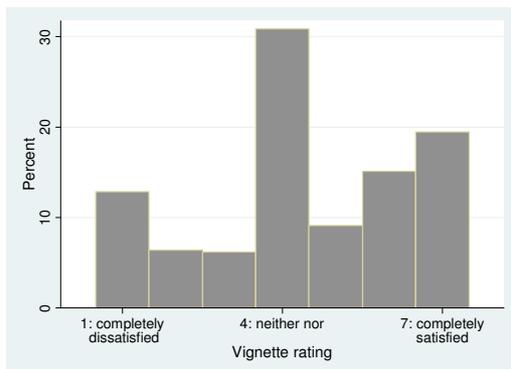
attention than the first (note that this module was included at the end of a very long set of questions on satisfaction with other facets of life).

Figure 5: Vignette ratings by order of vignette



In addition, there is very strong evidence of clumping at the midpoint among respondents who report the same rating for each of the three questions (see Figure 6), suggesting that these respondents have paid considerably less attention to the questions than those who gave different answers.

Figure 6: Vignette ratings for respondents using the same rating for all three vignettes



This indicates that a degree of caution may be required in using these responses. One solution may be to drop those respondents who give identical answers to all three vignettes, although this would involve working with a considerably smaller sample, and losing some valid responses (i.e., where the respondent was genuinely indifferent between scenarios). Preliminary investigations (see next section) indicate that dropping these respondents does not change our estimates much, and the precision of estimates remains about the same - the sample becomes smaller, but the variability of the remaining responses is larger.

Preliminary multivariate analysis

We performed preliminary analysis of the determinants of satisfaction with housework arrangements. Our findings include the following:

- Earning twice as much as your partner is associated with an increased preference for doing less than half the housework.

- By contrast, earning only half as much as your partner is *not* associated with an increased preference for doing more than half the housework!
- Having a young baby is associated with an increased preference for an equal allocation of housework.
- Having a teenage child is associated with a preference for an equal allocation – even more strongly than for a baby.
- In contrast to what one might expect, there are few statistically significant differences between men’s and women’s responses. This means that given the same earning power and hours of paid work, men and women express almost identical preferences over the distribution of housework. Thus, it appears that gender differences over the allocation of housework do not arise because of differences in preferences, and that we should look elsewhere for explanations for gender differences in the distribution of housework.

We also performed preliminary sensitivity analyses on these findings. The findings do not change if (a) we estimate an ordered logit model rather than a linear regression; (b) we estimate regressions separately for those under and over the age of 40, and for those with and without children.

We then checked whether personal characteristics affect the level of satisfaction which people report for the different scenarios. They do: among other findings, actual satisfaction with housework share is strongly predictive of satisfaction with the hypothetical scenarios, and (probably unsurprisingly) people who do a high share of the housework in real life are more likely to report being satisfied with a scenario where they do more than half the housework.

i. Family, Friends and Fertility: Context effects on childbearing intentions and close social network reporting (Paul Mathews).

Preceding questions can influence respondents' answers to later questions. These are known as ‘context effects’ as the context of the survey influences the responses of participants (Tourangeau, Singer et al. 2003, Rimal and Real 2005). Preceding questions have been shown to significantly influence the reporting of subjects as diverse as visual impairments (Todorov 2000), life satisfaction (Schwarz, Strack et al. 1991) and the approval of census data collection (Tourangeau, Singer et al. 2003). Much of this work has been conducted using convenience samples but their general frequency of context effect is argued to be fairly limited (Smith 1988). However, context effects are particularly worrying for longitudinal multi-purpose surveys, as observed changes could be the result of different preceding question batteries rather than being genuine changes over time. The Innovation Panel has already been used to investigate the risk of context effects on measuring political engagement (Uhrig 2012).

Here we look for a new set of potential context effects; an individual's opinions toward their future fertility¹⁰ and their close social network. Numerous factors such as partnership, age, actualised childbearing, economic position and social pressures predict the number of children that individuals report wanting or expecting (for examples see (Freedman, Hermalin et al. 1975, Schoen, Astone et al. 1999, Smallwood and Jefferies 2003, Berrington 2004, Testa and Toulemon 2006, Heiland, Prskawetz et al. 2008, Gipson and Hindin 2009, Kodzi, Casterline et al. 2010, Nettle, Coall et al. 2010, Iacovou and Tavares 2011). Because fertility intentions depend on so many factors, they are often subjective to sizeable uncertainty and thus malleability. Demographers therefore have longstanding concerns about the measurement and utility of such fertility attitudes because this uncertain (Westoff, Mishler et al. 1957, Ni Bhrolchain, Beaujouan et al. 2010). Previous work has shown that responses to fertility intention questions can be altered by preceding questions on mortality, particularly for males (Mathews and Sear 2008, Mathews 2012), which we have argued supports the notion that fertility intention are highly uncertain and malleable. We are not aware of any previous experiments looking at question ordering on individual's responses to questions on their social networks.

Methods

Our experiment ran across waves 4 and 5 of the Innovation Panel. At the household level participants were randomly allocated to one of two conditions. Approximately half of the participants were asked the questions on their fertility intentions *before* the measurement of their close social network and the other half were asked the questions on fertility preferences *after* their close social network questions. The specific fertility intentions questions were '*Do you think you will have any (more) children?*' If the participant answered '*Yes*' they were then asked '*How many (more) children do you think you will have?*' We focus analysis on the first question. The social network battery asked nine questions about the three individuals who live outside the household but are most emotionally proximate to the respondent.¹¹ Previously work has shown the number of relatives in this network predicts actual childbearing (Mathews and Sear 2013a, Mathews and Sear 2013b).

Our hypotheses were:

- Respondents who received the questions on fertility intentions *after* the questions on close family and friends will increase their reported fertility intentions. Making respondents think about their close social network (particular their relatives) would prime them to consider individuals who are likely to be most supportive of childbearing and most likely to provide informal 'free' childcare (Turke 1989; Newson, Postmes et al. 2005). The priming is effectively for lower cost children and thus should lead respondents to report stronger fertility intentions

¹⁰ We will throughout use the term fertility in the demographic sense, referring to actual childbearing. This should not be confused with fecundity, which is the biological capacity to become pregnant.

¹¹ Details of the exact question wording can be found <https://www.understandingsociety.ac.uk/about/innovation-panel>

○ Secondly, we predicted that the priming may work in the reverse direction. Respondents who received the questions on fertility intentions *before* the questions on close family and friends would be primed with thoughts of children and childcare. Therefore they may report more relatives (who are more liable to provide childcare) within their three closest individuals.

There was not a pure ‘control’ group as the questions before our experiment were different. In wave 4 the experiment was preceded by a question on the strength of the respondent’s opinions. Specifically they were asked the frequency which the following statement applied to them ‘*I’ve been able to make up my own mind about things*’ with the following answers available: 1 *None of the time*, 2 *Rarely*, 3 *Some of the time*, 4 *Often*, 5 *All of the time*. In wave 5 the preceding question before the experiment was on general happiness. Specifically the wording was ‘*Have you recently been feeling reasonably happy, all things considered?*’ Answer categories were: 1 *More so than usual*, 2 *About the same as usual*, 3 *Less so than usual*, 4 *Much less than usual*. These preceding questions are clearly different, though we did not have any a priori reason for predicting that they would influence the reporting of either fertility intentions or the close social network.

We restricted our analysis to individuals who had a valid response to the fertility intentions question i.e. we excluded the small number participants who refused to answer the question or said they were currently pregnant. We then ran bivariate analysis and also multivariate regression logistic models controlling for repeated measures and socio-economic background, predicting whether i) a respondent reported expecting a(nother) child or ii) had a relative in their close social network. The experiments were embedded within the CASI (computer assisted self-interviewing) section of the Innovation Panel and all analysis was conducted using Stata using the `xtmelogit` command to control for repeated measures from the same individuals in the multivariate models.

Results

In total there were 696 valid participants. Of these 226 individuals took part in both waves, 406 only took part in wave 4 and 64 only took part in wave 5. Our participants came from 281 households in wave 4 and 200 in wave 5.

Fertility intentions

Our results surprisingly suggest that if there was an outlier group, then it would be the control group in wave 4 who had a preceding question on ‘making up your own mind.’ This question would appear to potentially decrease the proportion of participants expecting a(nother) child. The simple bivariate results are set out in Table 1. The regression coefficients, controlling for various background features and repeated measurement are set out in Table 2. We also checked the effect within particular sub groups. We found that unmarried individuals seemed to be the most influenced by the ‘making up their mind’ preceding question, with a marginally statistically significant decrease in the probability of expecting a(nother) child.

Table 22: Percentage reporting expecting a(nother) child by wave and preceding question

	All participants		Just unmarried participants	
	Wave 4	Wave 5	Wave 4	Wave 5
Treatment – preceding close social network	34.4%	32.7%	50%	48.8%
Control 1 – make mind up question	27.0%	/	37.1%	/
Control 2 – general happiness question	/	33.6%	/	47.4%
Number of participants	409	287	225	160
P-value of a t-test between control and treatment within the wave (Note: without repeated measures correction)	0.052	0.44	0.03	0.43

Table 23: Regression coefficients with fixed effects for the individual predicting participant reporting expecting a(nother) child

	All participants		Just unmarried participants	
	Coef	p	Coef	p
Treatment – preceding close social network (ref)	/	/	/	/
Control 1 – make mind up question	-0.65	0.12	-0.79	0.07
Control 2 – general happiness question	-0.01	0.98	-0.13	0.79

Controlling for repeated measures within individuals, sex, age, has children, employed, household composition and marriage (in the all participants model)

It is difficult to explain this result. It does not particularly support our specific hypothesis, as the questions on happiness seemed to be just as pro-natal as the treatment. However, it does support the general notion the fertility intentions are volatile and difficult to measure.

Close social network

We did not find any evidence that respondents choose more relatives in their social network after being priming by questions on future fertility. The simple bivariate analysis is set out in Table 24 and multivariate regression coefficients in Table 25. We did not find any sizeable effects even confining analysis to particular sub-groups. We also investigated numerous other components of the social network composition (age, sex, time known, frequency of contact, geographic distance) but found very limited evidence for a context effect on the measure of these aspects of the network.

Table 24: Percentage reporting a relative in their close social network by wave and preceding question

	All participants	
	Wave 4	Wave 5
Treatment – preceding fertility intentions question	31.4%	26.5%
Control 1 – make mind up question	29%	/
Control 2 – general happiness question	/	25%
Number of participants	409	287
P-value of a t-test between control and treatment within the wave (Note: without repeated measures correction)	0.30	0.38

Table 25: Regression coefficients with repeated measurement control predicting participant reporting expecting a(nother) child (p value in brackets)

	All participants	
	Coef	p
Treatment – preceding fertility intentions question (ref)	/	/
Control 1 – make mind up question	0.36	0.27
Control 2 – general happiness question	-0.1	0.78

Controlling for repeated measures within individuals, sex, age, has children, employed, household composition and marriage (in the all participants model)

Conclusions

Given that we were only changing the preceding question we would expect relatively small effect sizes. We found a small but marginal significant effect for the fertility intentions question. This is the third time a preceding question priming effect has been observed on fertility attitudes (the others being Mathews and Sear (2008), Mathews (2012)), but with a new preceding question stimulus. This provides some further tentative evidence that fertility intentions are at risk of preceding question effects and reinforces the point about such attitudes being difficult to measure. Furthermore it suggests that there could be a substantial range of preceding questions that might alter later fertility intentions. The absence of an effect on the social network questions suggests that these items are less likely to be influenced by preceding questions.

However, our results, and any conclusions drawn from them, must be treated with caution. After multivariate controls were included the effect was only marginal in a particular subgroup. Moreover, the results were unpredicted and quite frankly surprising. We can only make post-hoc speculation for the pattern of results. It is plausible that unmarried individuals have greater uncertainty in their future childbearing attitudes, due to less certainty in future partnerships and the partners' opinions. This might make them more susceptible to priming effects.

The direction of the effect is harder to explain. The contemporary UK has high parental investment per child, low fertility and universal access to modern contraceptives. So childbearing often requires an active choice, which is made when the circumstances are 'right' to have a child. The wave 4 control question primed individuals to think about the frequency they were unable to make up their mind. If an individual is unable to make up their mind then it is plausible that the 'default' in such conditions is to *not* to expect to have a child; hence the anti-childbearing priming effect. Respondents also only had two answers to the question of expecting more children 'yes' or 'no.' The negative response category is likely contain within it a range of opinions and 'no' might well mean 'not now.' However, such an interpretation is far from certain. This interpretation of priming effects being focused on the immediate priming questions. It is plausible that the effects were actually driven by earlier more remote items.

The study provides a nice justification for repeated measurement in randomised experiments. If we had only had the first wave of data (i.e. wave 4) we would have naively and incorrectly concluded that the difference between treatment and the control was due to an active treatment, which was in the direction we had predicted. In fact it appears that the active element was the first control group, and the treatment and second control group were the inactive elements.

REFERENCES

- Bailar, B. A. (1975). "The effects of rotation group bias in estimates from panel surveys." *Journal of the American Statistical Association*. **70**: 23-30.
- Bartels, L. M. (1999). "Panel Effects in the American National Election Studies." *Political Analysis* **8**(1): 1-20.
- Battaglia, M. P., E. R. Zell, et al. (1996). "Can participating in a panel sample introduce bias into trend estimates?" 1996 Proceedings of Survey Research Methods Section fo the American Statistical Association: 1010-1013.
- Bem, Daryl J. 1972. "Self-Perception Theory." In Leonard Berkowitz (eds.), Advances in Experimental Social Psychology, vol. 6, pp. 1-62. New York: Academic Press.
- Berrington, A. (2004). "Perpetual postponers? Women's, men's and couple's fertility intentions and subsequent fertility behaviour." *Population Trends* **117**: 9-19.
- Blanchflower, D. G. and Oswald, A.J. (2004). "Well-being over time in Britain and the USA," *Journal of Public Economics*, Elsevier, vol. 88(7-8), pages 1359-1386, July.
- Blanden, J., and S. Machin (2004): "Educational Inequality and the Expansion of UK Higher Education," *Scottish Journal of Political Economy*, 51(2), 230-249.
- Budd, S., Gilbert, E., Butron, J., Jackle, A., Kaminska, O., Uhrig, SC., Brown, M., and Calderwood, L. "Understanding Society Innovation Panel Wave 4: Results from Methodological Experiments." Understanding Society Working paper Series 2012-06.
- Burton et al. (2012) Understanding Society Innovation Panel Wave 4: results from methodological experiments, Understanding Society Working Paper 2012-06, Colchester: University of Essex.
- Burton, J. (2011) "Understanding Society Innovation Panel Wave 3: Results from methodological experiments" Understanding Society Working Paper 2011-05, Colchester: University of Essex.
- Cantor, D. (1989). Substantive Implications of Longitudinal Design Features: The National Crime Survey as a Case Study. Panel Surveys. D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh. New York John Wiley: 25-51.
- Cantor, D. (2008). A Review and Summary of Studies on Panel Conditioning. Handbook of Longitudinal Research: Design, Measurement, and Analysis. S. Menard. London, Academic Press: 123-138.
- Clark A. and Oswald, A.J. (1996) "Satisfaction and comparison income" *Journal of Public Economics* 61 (1996) 359-381
- Clark, A. (1997) "Job satisfaction and gender: Why are women so happy at work?" *Labour economics*. Volume 4, Issue 4, December 1997, Pages 341–372

Clausen, A. (1968). "Response validity: Vote report." *Public Opinion Quarterly* **32**: 588-606.

Cohen, S. B. and V. L. Burt (1985). "Data collection frequency effects in the National Medical Care Utilization and Expenditure Survey." *Journal of Economic and Social Measurement* **13**: 125-151.

Easterlin, R.A. (1974). "Does Economic Growth Improve the Human Lot? Some Empirical Evidence", in Nations and Households in Economic Growth: Essays in Honour of Moses Abramowitz, edited by P.A. David and M.W. Reder, Academic Press, New York and London.

Easterlin, R.A. (1995). "Will Raising the Incomes of All Increase the Happiness of All?", *Journal of Economic Behavior and Organization*, *27*, 35-47.

Ferber, R. (1953). "Observations on a consumer panel operation." *Journal of Marketing* **17**: 246-259.

Freedman, R., A. I. Hermalin and M.-C. Chang (1975). "Do Statements About Desired Family Size Predict Fertility? The Case of Taiwan, 1967-1970." *Demography* **12**(3): 407-416.

Ghangurde, P. D. (1982). "Rotation group bias in the LFS estimates." *Survey Methodology* **8**: 86-101.

Gipson, J. D. and M. J. Hindin (2009). "The effect of husbands' and wives' fertility preferences on the likelihood of a subsequent pregnancy, Bangladesh 1998-2003." *Population Studies: A Journal of Demography* **63**(2): 135 - 146.

Gorassini, D.R. and Olson, J.M. 1995. "Does Self-perception Change Explain the Foot-in-the-Door Effect?" *Journal of Personality and Social Psychology* (69):91-105.

Groves, R. M., Cialdini, R. B., and Couper, M. P. "Understanding the Decision to Participate in a Survey." *Public Opinion Quarterly*, 1992, *56*, 475-495.

Heiland, F., A. Prskawetz and W. Sanderson (2008). "Are Individuals' Desired Family Sizes Stable? Evidence from West German Panel Data." *European Journal of Population/Revue européenne de Démographie* **24**(2): 129-156.

Iacovou, M. and L. P. Tavares (2011). "Yearning, Learning, and Conceding: Reasons Men and Women Change Their Childbearing Intentions." *Population and Development Review* **37**(1): 89-123.

Kemsley, W. F. F. (1961). "The household Expenditure Enquiry of the Ministry of Labor." *Applied Statistics* **10**(3): 117-135.

Kochanek, Kymn, Parvati Krishnamurty, Robert Michael, "The 'Early Bird' Field Innovation on a 30-year-old Longitudinal Survey". Paper presented at the 2nd Panel Survey Methods Workshop, Mannheim, Germany (5th-6th July 2010).

Kodzi, I., A. Kodzi, J. Casterline, B. and P. Aglobitse (2010). "The Time Dynamics of Individual Fertility Preferences Among Rural Ghanaian Women." *Studies in Family Planning* **41**(1): 45-54.

Kraut, R. E. and J. G. McConahay (1973). "How being interviewed affects voting: An experiment." *Public Opinion Quarterly* **37**(3): 398-406.

Krueger, A. (2007) "Are we Having fun Yet?" Categorizing and Evaluating Changes in Time Allocation." *Booking Papers on Economic Activity*, 2007.

Manski, C. (2004) "Measuring Expectations", *Econometrica*, Vol. 72, 5.

Mathews, P. (2012). Plasticity, life history and inclusive fitness: an evolutionary demography perspective on individual variation in fertility and fertility preferences in contemporary Britain. PhD, London School of Economics and Political Science.

Mathews, P. and R. Sear (2008). "Life after death: An investigation into how mortality perceptions influence fertility preferences using evidence from an internet-based experiment " *Journal of Evolutionary Psychology* **6**(3): 155-172.

Mathews, P. and R. Sear (2013). "Does the Kin Orientation of a British Female's Social Network influence her Entry into Motherhood?" *Demographic Research* **28**: 313-340.

Mathews, P. and R. Sear (2013). "Family and Fertility: Kin Influence on the Progression to a Second Birth in the British Household Panel Study." *PLoS ONE* **8**(3).

Mooney, H. (1962). Methodology in two California health surveys. US Public Health Service, Public Health Monograph No. 70. Washington, DC, US Government Printing Office.

Neter, J. and J. Waksberg (1964). "Conditioning effects from repeated household interviews." *Journal of Marketing* **28**: 51-56.

Nettle, D., D. A. Coall and T. Dickins, E. (2010). "Birthweight and paternal involvement predict early reproduction in British women: Evidence from the National Child Development Study." *American Journal of Human Biology* **22**(2): 172 - 179.

Ni Bhrolchain, M., E. Beaujouan and A. Berrington (2010). "Stability and change in fertility intentions in Britain, 1991-2007." *Population Trends* **141**: 13-35.

Norton, E.C., H. Wang and C. Ai (2004). "Computing Interaction Effects and Standard Errors in Logit and Probit Models." *The Stata Journal* **4**:154-167.

Pascale, J. and Mayer, T. S. (2004) Exploring Confidentiality Issues related to Dependent Interviewing: preliminary Findings, *Journal of Official Statistics*, 20(2), pp. 357-377.

- Rimal, R. N. and K. Real (2005). "Assessing the Perceived Importance of Skin Cancer: How Question-Order Effects Are Influenced by Issue Involvement." *Health Education Behavior* **32**(3): 398-412.
- Sala, E., Uhrig, S.C.N. and Lynn, P. (2011) "It is time computers do clever things!": the impact of dependent interviewing on interviewer burden, *Field Methods*, 23(1), 3-23.
- Schoen, R., N. M. Astone, Y. J. Kim, C. A. Nathanson and M. F. Jason (1999). "Do Fertility Intentions Affect Fertility Behavior?" *Journal of Marriage and Family* **61**(3): 790-799.
- Schwarz, N., F. Strack and H.-P. Mai (1991). "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis." *Public Opinion Quarterly* **55**(1): 3-23.
- Silberstein, A. R. and C. A. Jacobs (1989). Symptoms of repeated interview effects in the Consumer Expenditure Interview Survey. *Panel Surveys*. D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh. New York, Wiley: 289-303.
- Smallwood, S. and J. Jefferies (2003). "Family building intentions in England and Wales: Trends, outcomes and interpretations." *Population Trends* **112**: 15-28.
- Smith, T. W. (1988). Ballot Position: An Analysis of Context Effects Related to Rotation Design. *GSS Methodological Report*. Chicago, NORC. **55**.
- Stevenson, B and Wolfers, J (2009). "The Paradox of Declining Female Happiness," *American Economic Journal: Economic Policy*, American Economic Association, vol. 1(2), pages 190-225, August.
- Strenta, A. and Dejong, W. 1981. "The Effect of a Prosocial Label on Helping Behavior." *Social Psychology Quarterly* (44):142-147.
- Sturgis, P., N. Allum, et al. (2009). Attitudes Over Time: The Psychology of Panel Conditioning. *Methodology of Longitudinal Surveys*. P. Lynn. New York, John Wiley & Sons, Ltd: 113-126.
- Testa, M. R. and L. Toulemon (2006). "Family Formation in France: Individual Preferences and Subsequent Outcomes." *Vienna Yearbook of Population Research*: 41-75.
- Todorov, A. (2000). "Context Effects in National Health Surveys: Effects of Preceding Questions on Reporting Serious Difficulty Seeing and Legal Blindness." *Public Opinion Quarterly* **64**(1): 65-76.
- Tourangeau, R., E. Singer and S. Presser (2003). "Context Effects in Attitude Surveys: Effects on Remote Items and Impact on Predictive Validity." *Sociological Methods Research* **31**(4): 486-513.
- Turner, R. (1961). "Inter-Week Variations in Expenditure Recorded During a Two-Week Survey of Family Expenditure." *Applied Statistics* **10**(3): 136-146.

Uhrig, N. (2012). IP4 Context effects. "Understanding Society Innovation Panel Wave 4: Results from Methodological Experiments". J. Burton, Understanding Society Working Paper Series. **6**.

Uhrig, S.C.N. (2012) "Understanding panel conditioning: An examination of social desirability bias in self-reported height and weight in panel surveys using experimental data". *Journal of Longitudinal and Life Course Studies*, 3:1, pp 120-136.

Waterton, J. and D. Lievesley (1989). Evidence of conditioning effects in the British Social Attitudes Panel. Panel Surveys. D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh. New York, John Wiley & Sons: 319-339.

Westoff, C. F., E. G. Mishler and E. L. Kelly (1957). "Preferences in Size of Family and Eventual Fertility Twenty Years After." *The American Journal of Sociology* **62**(5): 491-497.

Wilson, S. E. and B. L. Howell (2005). "Do panel surveys make people sick? US arthritis trends in the Health and Retirement Survey." *Social Science and Medicine* **60**(11): 2623-2627.

Yalch, R. F. (1976). "Pre-election interview effects on voter turnout." *Public Opinion Quarterly* **40**: 331-336.